2014

On Harmonic Vibrational Frequency Convergence to the Complete Basis Set Limit in Water Dimers and Hydrogen Fluoride Dimers

Amanda Hardwick
University of Mississippi. Sally McDonnell Barksdale Honors College

Examination of Harmonic Vibrational Frequency Convergence to the Complete Basis Set Limit in Water Dimers and Hydrogen Fluoride Dimers

By
Amanda Hardwick

A thesis submitted to the faculty of the University of Mississippi in partial fulfillment of the requirements of the Sally McDonnell Barksdale Honors College.

Oxford
May 2014

Approved by

_____
Advisor: Dr. Gregory Tschumper

_____
Reader: Dr. Susan Pedigo

_____
Reader: Dr. Nathan Hammer

For God who strengthens me, my family who supports me, and the university that challenges me.

Parents James and Leann Hardwick
Siblings Jaime and Mica
And best friends

ACKNOWLEDGEMENTS

Research for this thesis was performed under the supervision of Dr. Tschumper. Research was funded by EPSCoR grant EPS-0903787.

TABLE OF CONTENTS

1. ABSTRACT

This work focuses on the convergence of harmonic vibrational frequencies in the water dimer $(H_2O)_2$ and the hydrogen fluoride dimer $(HF)_2$ to the complete basis set (CBS) limit. It also evaluates the effect of counterpoise (CP) correction procedures and the effect of sequentially removing diffuse functions from specific basis sets on the harmonic vibrational frequencies of these dimers. Computations were run using second-order Møller-Plesset perturbation theory (MP2) with large, flexible correlation consistent basis sets. This research aims to determine if removing diffuse functions can increase the speed of harmonic vibrational frequency computations without sacrificing accuracy. Results indicate that CP-correction procedures are not cost-effective.

In order to get harmonic vibrational frequencies that deviate no more than 10 cm^{-1} from the CBS limit for these systems, a basis set of at least haQZ quality should be used. The average deviation in non-CP corrected frequencies using a TZ quality basis set from the CBS limit was 14 cm^{-1}. The average deviation in CP-corrected frequencies using a TZ quality basis set from the CBS limit was 17 cm^{-1}. In contrast, these deviations in non-CP corrected and CP-corrected frequencies using a QZ quality basis set from the CBS limit were both 7 $cm\text{-}^{1}$. CP-correction procedures are not cost-effective when basis sets of this size are used, however. Results also indicate that *f* and *g* diffuse functions can be removed from the oxygen

and fluorine atoms in haQZ and aQZ basis sets—and also *d* and *f* diffuse functions from hydrogen in the aQZ basis set—without appreciably increasing deviations from the CBS limit frequencies.

2. INTRODUCTION

2a. Non-covalent interactions

The vast majority of chemical properties and processes are largely determined by the behavior of electrons in molecules. Bonds between molecules are due to the tendency of electrons to be attracted to protons and repelled from other electrons. Non- covalent interactions are weaker than covalent interactions.[1, 2, 3] Covalent interactions are on the order of 100 kcal/mol while non-covalent interactions are on the order of 10 kcal/mol. Types of non-covalent interactions between neutral species include dispersion forces, dipole-dipole interactions, and hydrogen bonds.[4, 5, 6]

Dispersion forces are the weakest intermolecular force. Intramolecular forces are forces within a molecule, while intermolecular forces are forces between molecules. Dispersion forces are the temporary result of two adjacent molecules with uneven charge distributions where the high electron density end of one molecule becomes attracted to the low electron density end of the adjacent molecule. These dispersion forces are seen between all polar and non-polar molecules, and are stronger in molecules in which the valence electrons are farther away from the nuclei. Polarizablity is the measure of the ease with which the charge density of a molecule can be distorted. The polarization of a molecule leads to what is called an induced (or temporary) dipole. These induced dipoles can in turn induce

dipoles in adjacent molecules, creating an induced dipole-induced diploe interaction, a slightly stronger intermolecular interaction than dispersion forces.

When a highly electronegative atom such as oxygen, nitrogen, or fluorine covalently binds to an electropositive atom such as hydrogen the result is a permanent uneven charge distribution, or a permanent dipole. Permanent dipoles can create an induced dipole in a dipole-induced dipole interaction or they may participate in a dipole-dipole interaction. Dipole-dipole interactions are the strongest of the non-covalent interactions between neutral species. Dipole-dipole interactions are intermolecular interactions that are the result of two molecules with permanent dipole moments (unevenly distributed charge density).

2b. Hydrogen bonds

Hydrogen bonding is a particularly strong type of dipole-dipole interaction.[7] A hydrogen bond is formed when a hydrogen atom covalently bonds with a highly electronegative atom such as nitrogen, oxygen, or fluorine—creating a substantial partial positive charge on the hydrogen atom that is attracted to a region of another molecule with a partial negative charge. A hydrogen bond typically has less than ten percent of the strength of a covalent bond, making them challenging to study, both experimentally and computationally.[8, 9, 10, 11, 12]

Non-covalent interactions are very important in a number of areas because they play an important role in many of the physical and biochemical structures and processes in our bodies.[4, 5, 13, 14, 6, 15, 16, 17] Computational chemists have made continual efforts to examine the

properties of the hydrogen bonds that play such a dramatic role in our bodies and in the water that covers our planet. One approach to studying hydrogen bonding that computational chemists employ is to create molecular models of very basic hydrogen bonds, for example the hydrogen fluoride dimer or water dimer.[5, 6] They can progressively model larger and larger molecule clusters to learn how the hydrogen bonding changes as the system size increases. Studying water dimer systems can give chemists an idea how strong hydrogen bonding is, which is important to enhance our knowledge of water. This research is primarily focused on the water and hydrogen fluoride dimer systems because they are fundamental models for hydrogen bonding.[18]

2c. Water

Water is one of the key components of life. Water makes up about seventy percent of the earth's surface and accounts for approximately sixty percent of the mass of an adult man. Water lubricates joints and absorbs shock to the brain and spinal cord. It is a main component of blood, making it a player in regulation of body temperature and oxygen transport as well as excretion.[19] Water is also the only substance on earth to occur naturally in the solid, liquid, and gas states.[19] Because it dissolves more substances than most other liquids, even some acids, water is known as the "universal solvent."[19] Water plays a role in making and breaking glycosidic linkages through condensation and hydrolysis. Another interesting property of water is its incredibly strong cohesive strength—it is more cohesive than most other nonmetal liquids.[19] Cohesion is the tendency of a substance to stick to itself,

where adhesion is the tendency of a substance to stick to something other than itself. Water's cohesive force is the reason that doing a belly flop hurts—the surface tension is due to the tendency of water to interact so strongly with itself. Cohesion is responsible for the formation of water droplets and the reason why if you add water drop by drop to a penny the water stays on the penny rather than running off.

Water's incredible cohesive properties are due to the strong hydrogen-bonding network that a water molecule forms with other water molecules. Because of water's very strong hydrogen bonds, the water dimer is an ideal model for computational chemists to study hydrogen bonding.[6, 20, 21, 22] The water dimer is also a fairly small, simple system so it is more quickly and easily studied using computational chemistry methods than a larger clusters or bulk-phase water.[23, 24, 25, 26, 27] Studying the water dimer and others like it is important for computational chemists because these interactions are still not completely understood.

2d. Computational background

i. Methods

To understand these subtle non-covalent interactions, chemists depend on reliable theoretical models, in particular, quantum mechanical wavefunction methods. Ab initio wavefunction methods offer the most reliable description of these important but relatively weak interactions. A wavefunction method specifies a particular strategy for obtaining approximate solutions to the non-relativistic time-independent Schrödinger equation within

the Born- Oppenheimer approximation. [28]

The time-independent Schrödinger equation is represented by $\hat{H}\Psi=E\Psi$, where $\hat{H}$ represents the Hamiltonian operator, E represents the energy of the system, and Ψ represents the wavefunction.[28] The foundation of these techniques is typically the Hartree-Fock (HF) method, which treats an electron as if it feels an "average" field of the other electrons. However, the motions of individual electrons are correlated with each other. The Hartree-Fock approach only approximates the correlation of the electrons, which is why methods that improve on the HF model are often called correlated methods.[28] More sophisticated methods or models translate to longer calculation times and increased computational requirements. Therefore it is crucial to determine whether or not the additional accuracy is worth the additional computer time. The correlated method used in this research is the second order Møller-Plesset perturbation theory (MP2) method.[29]

ii. Basis sets

Because "Molecules are made up of atoms…molecular solutions [can] be made up of atomic solutions."[28] This linear combination of atomic orbitals (LCAO) approximation is then derived from the fact that "molecular orbitals (ψ) are expressed as linear combinations of a basis set of prescribed functions known as basis functions ϕ. The LCAO approximation is exact as the complete basis set limit is reached."[28, 30] The complete basis set (CBS) limit refers to the numbers acquired by using an infinitely large basis set. In this research, the CBS limit for the harmonic vibrational frequencies is estimated from computations with very large basis sets. The specific CBS limit for this research is described later.

"Basis sets are made up of a finite number of well-defined functions centered on each atom."[28] Gaussian-type orbital basis sets are widely used because the integrals can be greatly simplified, in comparison to the "Slater-type" functions, which are the solutions to the Schrödinger equation for the ground state of the hydrogen atom. Linear combinations of multiple Gaussian functions can be used to mimic the Slater-type functions. These linear combinations are called contracted functions, and they produce a better atomic orbital fit.[28]

For each system, there is a minimal basis set, or a minimum number of functions that is required to represent the chemical system. In a minimal basis set there is one basis function for each type of occupied subshell. For hydrogen, one function describing the 1s subshell is the minimal basis set. For oxygen, the minimal basis set is five functions: one to describe the 1s subshell, one to describe the 2s subshell, and three to describe the 2p subshell (i.e. $2p_x$, $2p_y$, and $2p_z$). Minimal basis sets are inaccurate, so more sophisticated basis sets have been developed and are more commonly used. One of these improved basis sets is the double-zeta (DZ) basis set. This basis set utilizes twice as many sets of functions to describe each occupied subshell. An oxygen atom would have a total of ten basis functions in the basis set: two describing the 1s orbital, two describing the 2s orbital, and six describing the 2p subshell. This procedure can be extended to create triple-zeta, quadruple-zeta, and even larger basis sets.

Polarization functions can also be added to basis sets to improve accuracy. Polarization functions are particularly important for accurately describing electron correlation. The simplest polarization function addition to a DZ basis set for water would

include a polarized *d* function on the oxygen and a *p* function on the hydrogen atom.[28] Higher angular momentum polarization functions can be added if desired.

This research utilizes correlation-consistent basis sets, which are designed specifically to converge to the CBS limit by systematically adding higher angular momentum polarization functions to each successively larger basis set. One example of these basis sets is the cc-pVQZ basis set. The –cc– denotes correlation-consistent and the p means that polarization functions have been added. The VQZ denotes that a quadruple-zeta basis set has been used for valence subshells. Correlation-consistent basis sets are useful in research because they are specifically designed to converge to the CBS limit.

For anions, atoms in excited states, and hydrogen bonding systems adding diffuse functions improves computation accuracy by accounting for the fact that the additional electron(s) may be bound more loosely to atoms.[28] These diffuse functions provide a clearer picture of interactions between molecules. Diffuse functions are useful in highly electronegative atoms such as those involved in hydrogen bonding because hydrogen bonds are long-range interactions. Diffuse functions are denoted in correlation-consistent basis sets by the prefix aug- meaning augmented. For example cc-pVQZ basis set with diffuse functions added becomes aug-cc-pVQZ. In such a basis set, diffuse functions are added to all atoms. In basis sets with diffuse functions added to only the "heavy" or non-hydrogen atoms, the term heavy is added before the augmented command. One of these basis sets would be called heavy-aug-cc-pVQZ (haQZ for short).

One final aspect of choosing a basis set to be aware of is that all finite basis sets, or

all those basis sets with a finite number of basis functions, for computing binding or interaction energies result in what is called a basis set superposition error (BSSE),[4, 7, 32, 33, 35] where the magnitude of the calculated interaction is artificially too large. The BSSE is only an issue in computations involving interaction or binding energies, and is closely related to the basis set incompleteness error (BSIE), which is also unavoidable when using finite basis set sets. The size of the BSSE decreases as the size of the basis set increases. Another method computational chemists utilize to try to eliminate the BSSE is the counterpoise-correction procedure.[1, 7]

A counterpoise-correction procedure works as follows. The typical equation for interaction energy is shown in Eq. 1.[31]

$$\mathrm{E_{int}(AB)} = \mathrm{E_{AB}^{AB}(AB)} - \mathrm{E_{AB}^{A}(A)} - \mathrm{E_{AB}^{B}(B)} \quad (1)$$

The basis set is denoted by the superscripts, the geometry is denoted by the subscripts, and the chemical system is denoted by the species in parentheses. Therefore, the interaction energy of dimer AB is equal to the energy of the dimer computed in the dimer basis set at the dimer geometry minus the energies of monomers A and B in their own basis sets and the geometries they adopt in the complex. This interaction energy is not corrected for the BSSE. Boys and Bernardi introduced a procedure called a counterpoise (CP) correction to help account for the BSSE by computing the energy of each monomer in the dimer basis set, which leads to the following expression for the CP-corrected interaction energy (Eq. 2).

$$\mathrm{E_{int}^{CP}(AB)} = \mathrm{E_{AB}^{AB}(AB)} - \mathrm{E_{AB}^{AB}(A)} - \mathrm{E_{AB}^{AB}(B)}$$

This corrected interaction energy is the energy of the dimer with dimer geometry and basis set minus the energy of monomer A with a dimer geometry and dimer basis set and monomer B with a dimer geometry and a dimer basis set.

CP-correction procedures have been shown to perform better than non-CP correction procedures under some circumstances and more poorly than the non-CP correction procedures in others.[32, 33, 34, 35, 36, 37, 38] Some argue that CP-correction procedures tend to "overcorrect" for the BSSE, especially in hydrogen-bonded complexes.[38] This research examines the effect of CP-corrections on harmonic vibrational frequencies. The goal of this work is to learn how rapidly the computed vibrational frequencies of the hydrogen-bonded systems converge to the CBS limit in order to gauge the reliability of results obtained with smaller basis sets and/or the CP-correction procedure.[39, 40, 41] Compact basis sets that can reproduce high-accuracy results are highly desirable because rigorous analyses with larger basis sets are prohibitively expensive.[39, 40, 41, 42, 43] More specifically, this work examines and compares the convergence of non-CP and CP corrected frequencies to the CBS limit for the $(H_2O)_2$ and $(HF)_2$ systems. This study also aims to determine if removing certain diffuse functions can reduce the time required to compute harmonic vibrational frequencies without sacrificing accuracy.

3. METHODS

The goal of this research was to determine the effect of CP-correction procedures and the removal of certain diffuse functions on the harmonic vibrational frequencies of $(H_2O)_2$ and $(HF)_2$. More specifically, it examined the difference between CP-corrected and non-CP corrected frequencies as they converge to the MP2 CBS limit. It also aimed to determine if the systematic removal of diffuse functions can help computational chemists reach the CBS limit more quickly in small hydrogen-bonded systems.[44, 45, 46, 47, 48] Full geometry optimizations and harmonic vibrational frequency computations were performed on the lowest energy configuration of $(H_2O)_2$ and $(HF)_2$ using the Gaussian09[49, 50] computational chemistry program. For each dimer system, both CP and non-CP corrected optimizations and frequencies were run in C_s symmetry using the MP2 method.

The basis sets used are all of the correlation-consistent (-cc-) variety, meaning that each basis set is designed to systematically converge to the CBS limit. The basis sets used in this research were XZ, haXZ, and aXZ, where XZ ranged from DZ to 6Z. The harmonic vibrational frequencies are compared to the MP2 results at the CBS limit. The CBS numbers are the averages of the harmonic vibrational frequency values at the CP-corrected and non-CP corrected a6Z and ha6Z basis set levels.[38]

To investigate the effects of the removal of certain diffuse functions on the harmonic vibrational frequencies, *g* diffuse functions are removed from the heavy atom in

the haQZ basis set for $(H_2O)_2$ and $(HF)_2$. In the aQZ basis set, *g* diffuse functions are removed from the heavy atom and *f* diffuse functions are removed from the hydrogen atom. The frequencies from the procedures where diffuse functions are systematically removed from non CP- corrected dimer basis sets are also compared to the non-CP corrected haQZ and aQZ frequencies as well as to the CBS limit. A detailed procedure for running the computations can be found in **Appendix 1A**.

4. RESULTS AND DISCUSSION

The hydrogen fluoride dimer

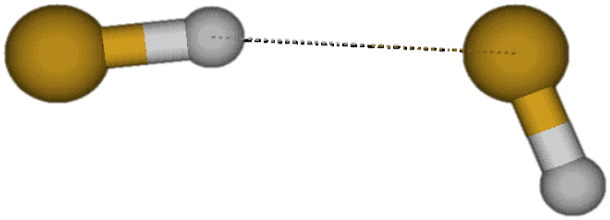

Fig. 1 The $(HF)_2$ system constructed for this research.

The water dimer

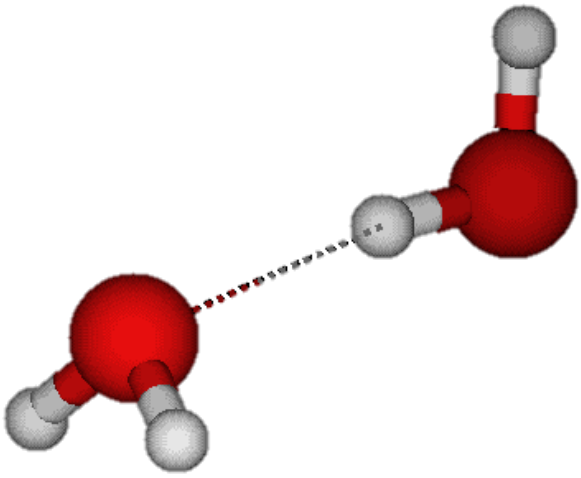

Fig. 2 The $(H_2O)_2$ system constructed for this research.

4a. The complete basis set limit

The estimated CBS frequencies for the hydrogen fluoride dimer are the average of the frequencies from the CP-corrected and non-CP corrected a6Z and ha6Z basis sets. The estimated CBS frequencies for the water dimer are the average of the frequencies from the CP-corrected and non-CP corrected ha6Z and the non-CP corrected a6Z basis sets. *Tables 1* and *2* show the frequencies from these basis sets and the estimated CBS frequencies for $(H_2O)_2$ and $(HF)_2$.

Table 1. The estimated CBS limit for $(HF)_2$

Non-CP corrected	ha6Z	a6Z	CP-corrected	ha6Z	a6Z	CBS estimate
	159	160		158	158	159
	216	218		215	215	216
	463	468		460	460	463
	567	569		563	563	566
	4001	4000		4002	4002	4001
	4094	4094		4094	4094	4094

Table 2. The estimated CBS limit for $(H_2O)_2$

Non-CP corrected	ha6Z	a6Z	CP-corrected	ha6Z	a6Z	CBS estimate
	122	121		121	----	121
	146	143		145	----	145
	150	150		149	----	150
	184	185		183	----	184
	352	352		350	----	351
	623	624		620	----	622
	1632	1632		1632	----	1632
	1651	1651		1651	----	1651
	3738	3737		3739	----	3738
	3834	3834		3835	----	3834
	3937	3937		3937	----	3937
	3957	3956		3957	----	3957

4b. The effects of counterpoise-correction procedures

The results of the harmonic vibrational frequency computations suggest that CP-correction procedures are not beneficial for converging to the CBS limit in small, hydrogen-bonded systems. When the harmonic vibrational frequencies without the CP-correction procedure (*Fig. 3*) are compared to those with CP-correction procedures (*Fig. 4*), there is no clear improvement over the non-CP corrected frequencies.

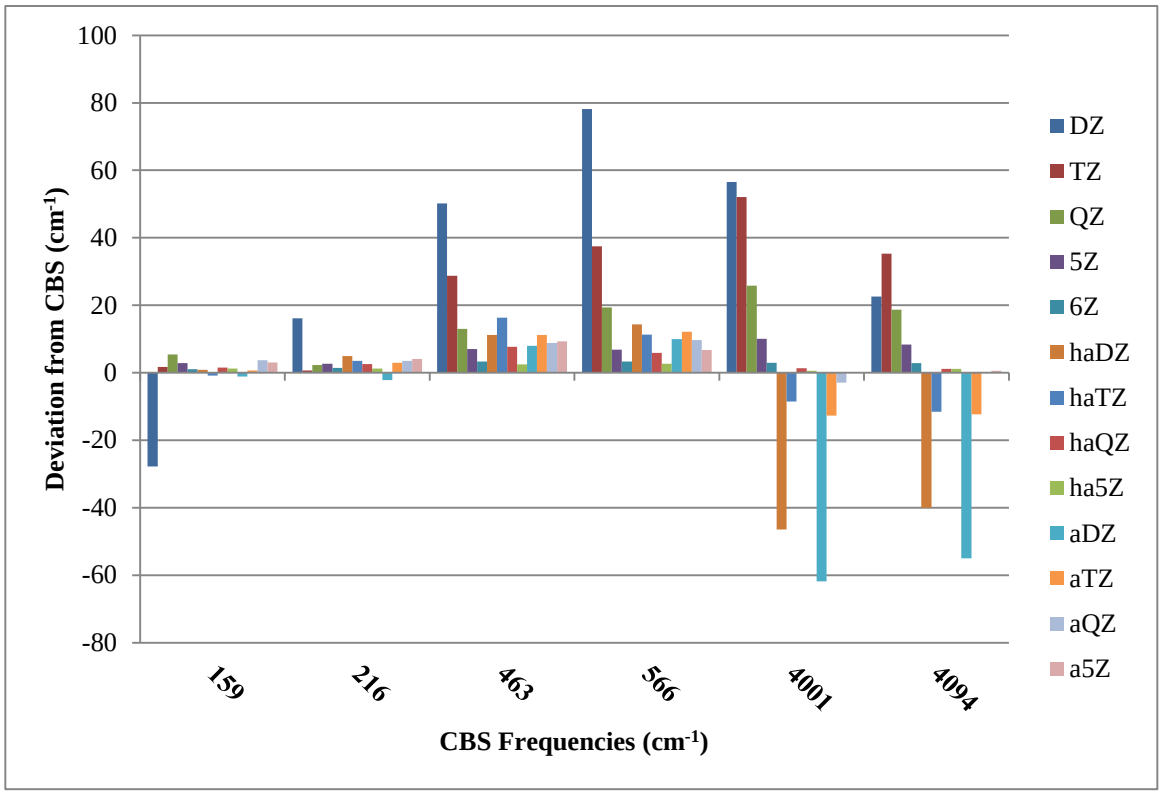

Fig. 3 Deviations from the estimated CBS limit harmonic vibrational frequencies without CP-correction procedures in $(HF)_2$

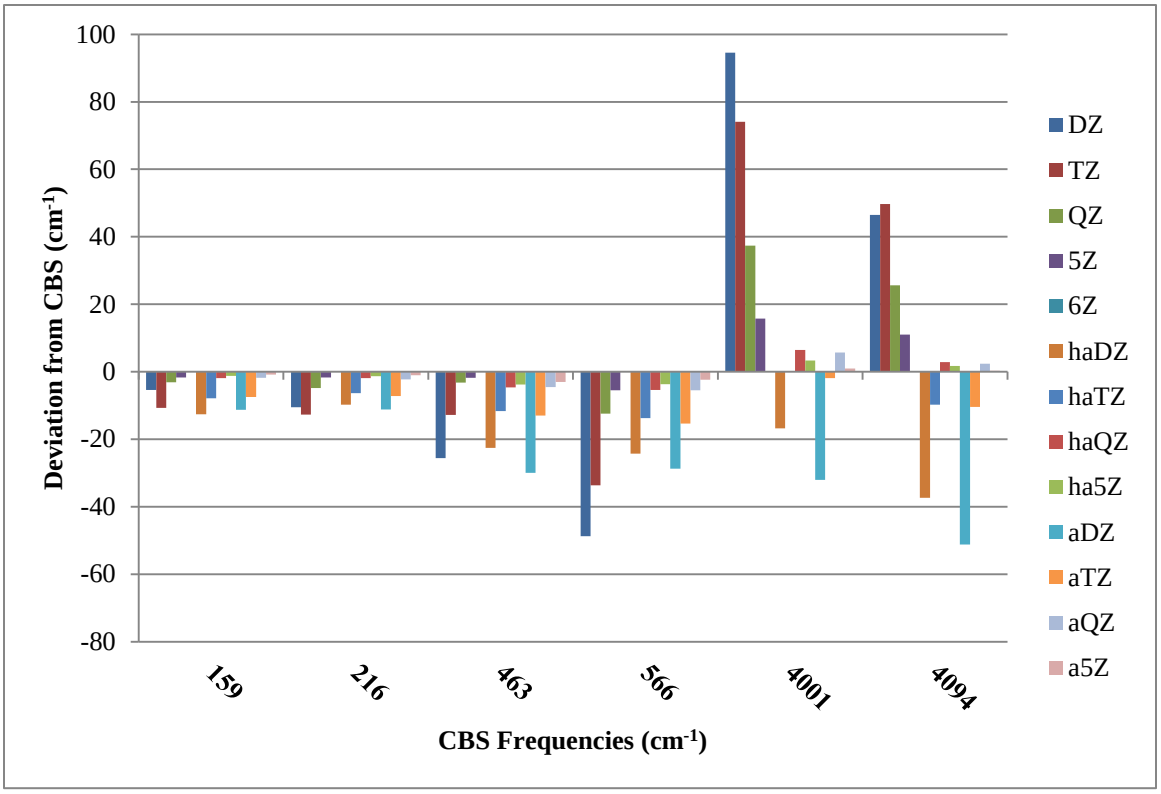

Fig. 4 Deviations from the estimated CBS limit harmonic vibrational frequencies with CP-correction procedures in $(HF)_2$

The differences in the deviations from the estimated MP2 CBS limit in the non-CP corrected and CP-corrected harmonic vibrational frequencies can be seen in *Figure 5-7*. *Figure 5* shows the deviations from the CBS estimate in the DZ, haDZ, and aDZ basis sets with and without CP-correction procedures. The maximum absolute deviation from the CBS estimate in the non-CP corrected frequencies is 78 cm^{-1} and the maximum absolute deviation from the CBS estimate in the CP corrected frequencies is 95 cm^{-1} with the DZ basis set. With the haDZ basis set, the maximum absolute deviations shrink to 46 cm^{-1} in the non-CP corrected frequencies and 37 cm^{-1} in the CP-corrected frequencies. With the aDZ basis set, the maximum absolute deviations shrink to 62 cm^{-1} in the non-CP corrected

frequencies and 51 cm^{-1} in the CP-corrected frequencies. These results show that when using DZ quality basis sets, CP-correction procedures do appear somewhat beneficial, but neither the CP-corrected or non-CP corrected frequencies are converged to within 10 cm^{-1} of the estimated MP2 CBS limit.

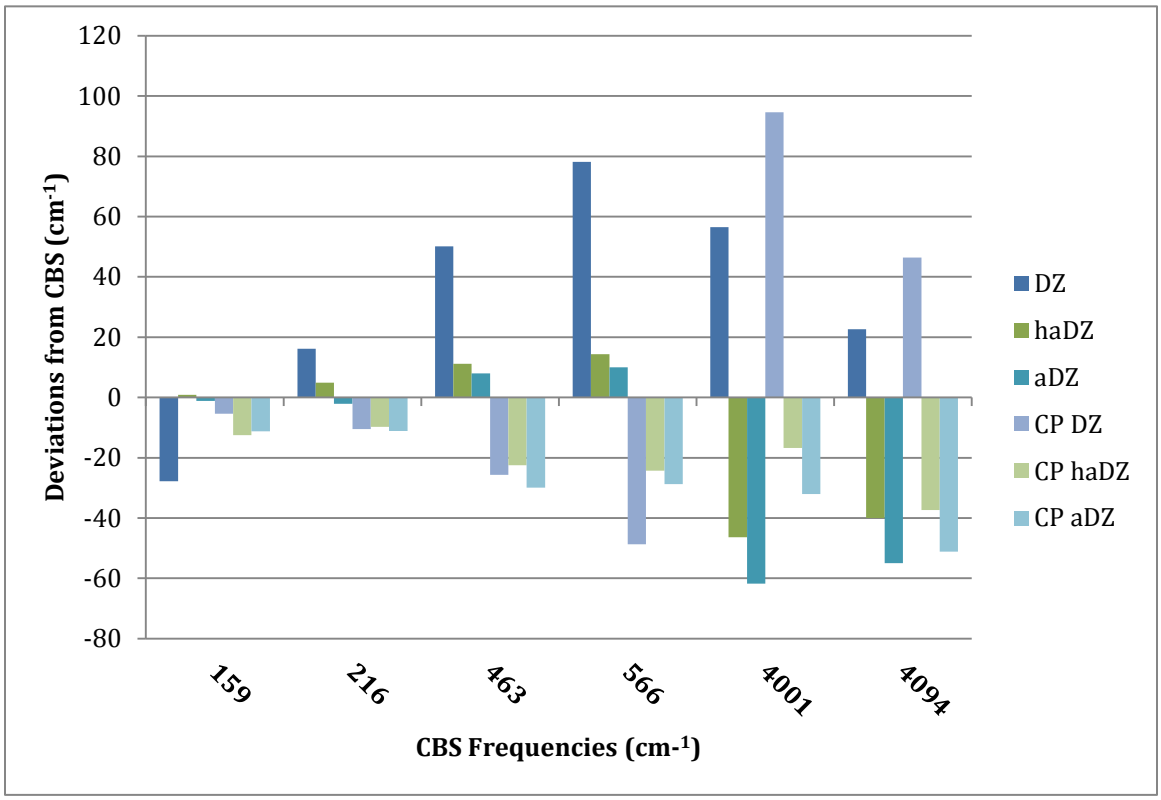

Fig. 5 Deviations from the estimated CBS limit harmonic vibrational frequencies in DZ quality basis sets with and without CP-correction procedures in $(HF)_2$

The deviations from the CBS estimate in the TZ, haTZ, and aTZ basis sets with and without CP-correction procedures are shown in *Figure 6*. The maximum absolute deviation from the CBS estimate in the non-CP corrected frequencies is 52 cm^{-1} and the maximum absolute deviation from the CBS estimate in the CP corrected frequencies is 74 cm^{-1} with the TZ basis set. With the haTZ basis set, the maximum absolute deviations shrink to 16 cm^{-1} in the non-CP corrected frequencies and 14 cm^{-1}

in the CP-corrected frequencies. With the aTZ basis set, the maximum absolute deviations shrink to 13 cm^{-1} in the non-CP corrected frequencies and 15 cm^{-1} in the CP-corrected frequencies. When the size of the basis set is increased to the TZ quality level, CP-corrections appear beneficial only when the haTZ basis set is used—and only decrease the deviation by 2 cm^{-1} over the non-CP corrected frequency deviations.

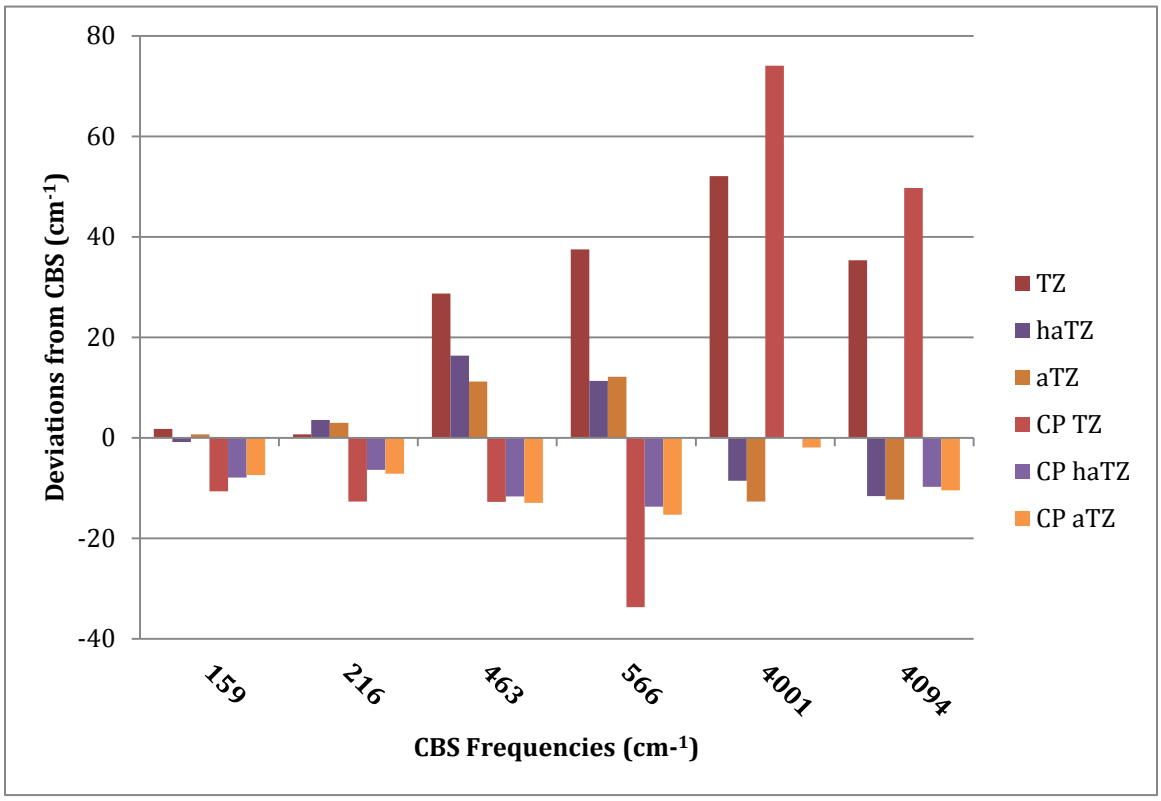

Fig. 6 Deviations from the estimated CBS limit harmonic vibrational frequencies in TZ quality basis sets with and without CP-correction procedures in $(HF)_2$

Figure 7 shows the deviations from the CBS estimate in the QZ, haQZ, and aQZ basis sets with and without CP-correction procedures. The maximum absolute deviation from the CBS estimate in the non-CP corrected frequencies is 26 cm^{-1} and the maximum absolute deviation from the CBS estimate in the CP corrected frequencies is 37 cm^{-1} with the QZ basis set. With the haQZ basis set, the maximum

absolute deviations shrink to 5 cm^{-1} in the non-CP corrected frequencies and 6 cm^{-1} in the CP-corrected frequencies. With the aQZ basis set, the maximum absolute deviations shrink to 10 cm^{-1} in the non-CP corrected frequencies and 6 cm^{-1} in the CP-corrected frequencies. These results indicate that in order to provide deviations from the CBS estimate of less than 10 cm^{-1}, a basis set of haQZ or larger needs to be utilized. These results also indicate that once a basis set of this size is used, the CP-corrections are not cost effective as the deviations in harmonic vibrational frequencies are already small.

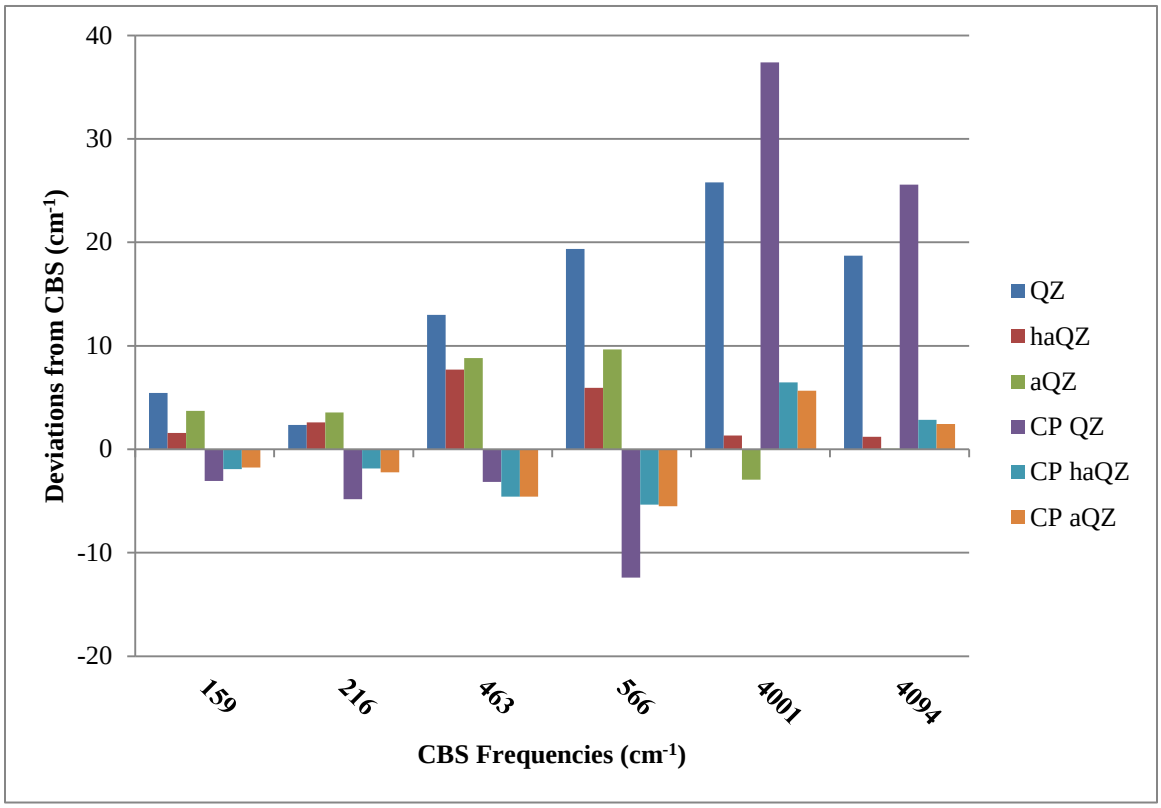

Fig. 7 Deviations from the estimated CBS limit harmonic vibrational frequencies in QZ quality basis sets with and without CP-correction procedures in $(HF)_2$

To further investigate the effect of CP-correction procedures on the deviations in harmonic vibrational frequencies from the estimated MP2 CBS limit further, the

deviations from the estimated CBS in non-CP corrected harmonic vibrational frequencies (*Fig. 8*) were compared to the CP-corrected harmonic vibrational frequencies (*Fig. 9*). Looking at the two sets of frequency deviations, the CP-corrected frequencies show no clear improvement.

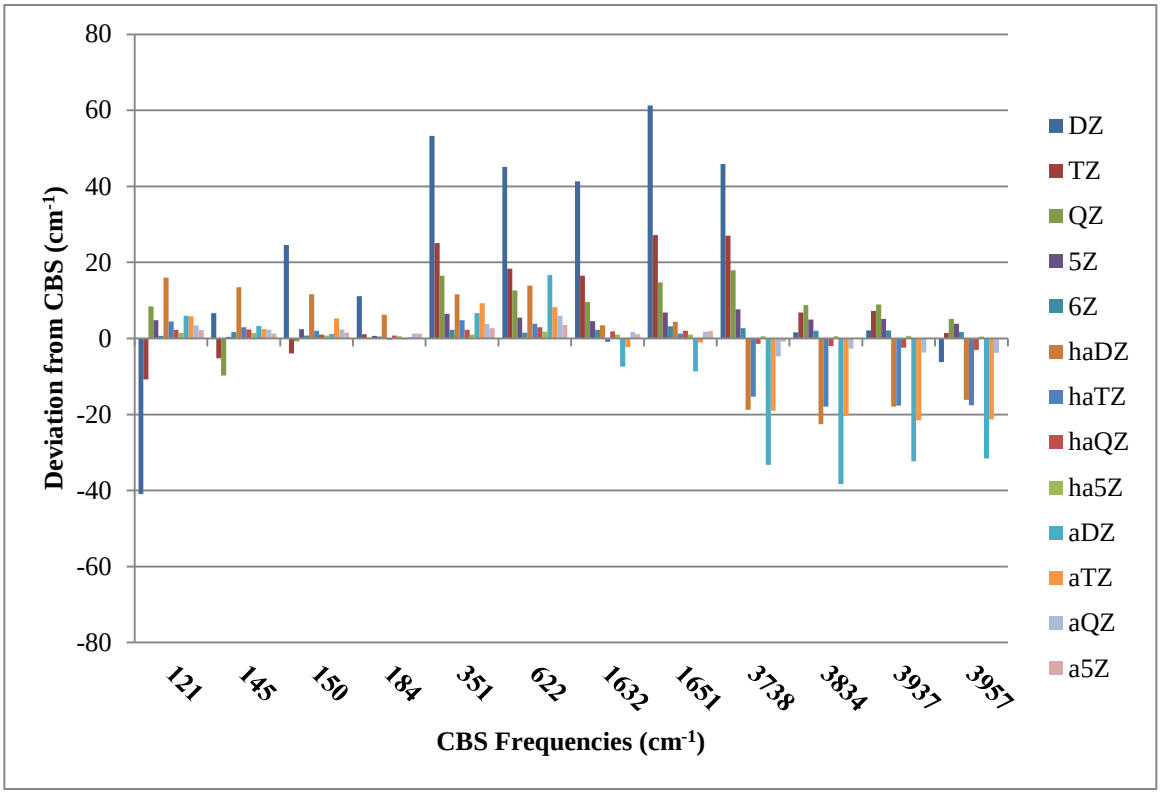

Fig. 8 Deviations from the estimated CBS limit harmonic vibrational frequencies without CP-correction procedures in $(H_2O)_2$

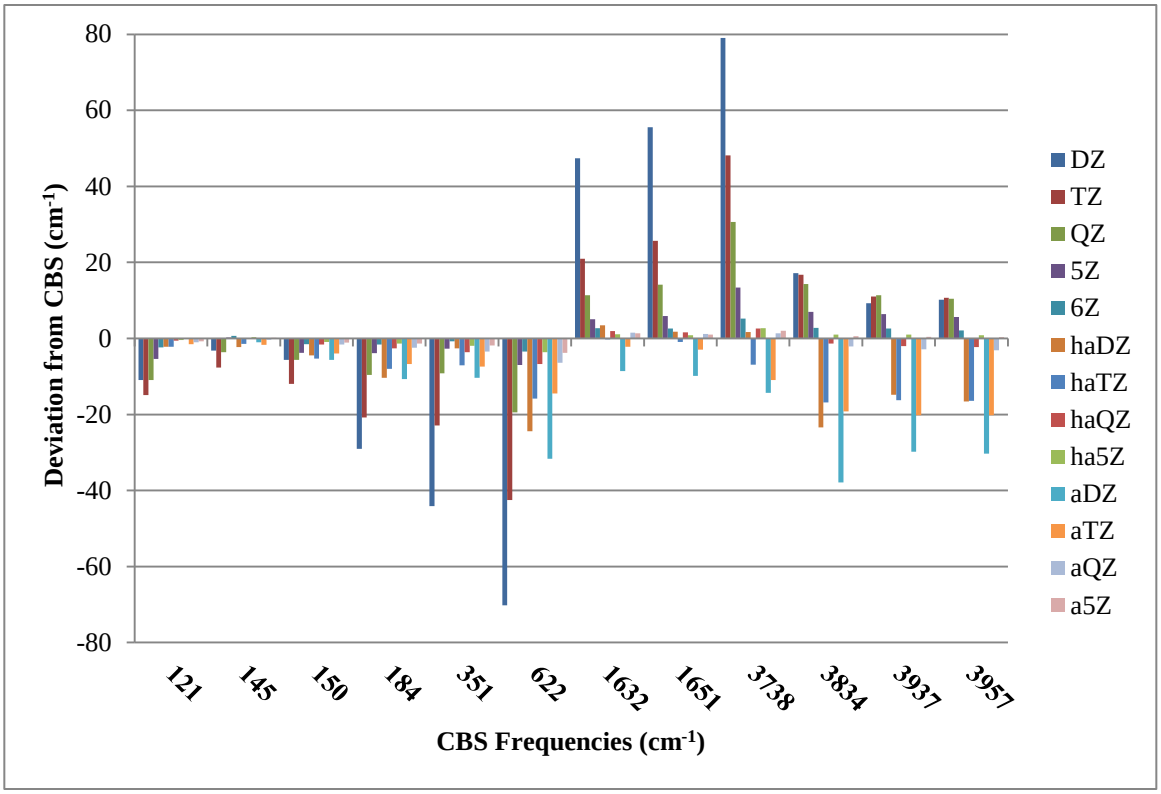

Fig. 9 Deviations from the estimated CBS limit harmonic vibrational frequencies with CP-correction procedures in $(H_2O)_2$

Figure 10 shows the deviations from the CBS estimate in the DZ, haDZ, and aDZ basis sets with and without CP-correction procedures in the water dimer. The maximum absolute deviation from the CBS estimate in the non-CP corrected frequencies is 61 cm^{-1} and the maximum absolute deviation from the CBS estimate in the CP corrected frequencies is 79 cm^{-1} with the DZ basis set. With the haDZ basis set, the maximum absolute deviations shrink to 23 cm^{-1} in the non-CP corrected frequencies and 24 cm^{-1} in the CP-corrected frequencies. With the aDZ basis set, the maximum absolute deviations shrink to 38 cm^{-1} in the non-CP corrected frequencies and 38 cm^{-1} in the CP-corrected frequencies.

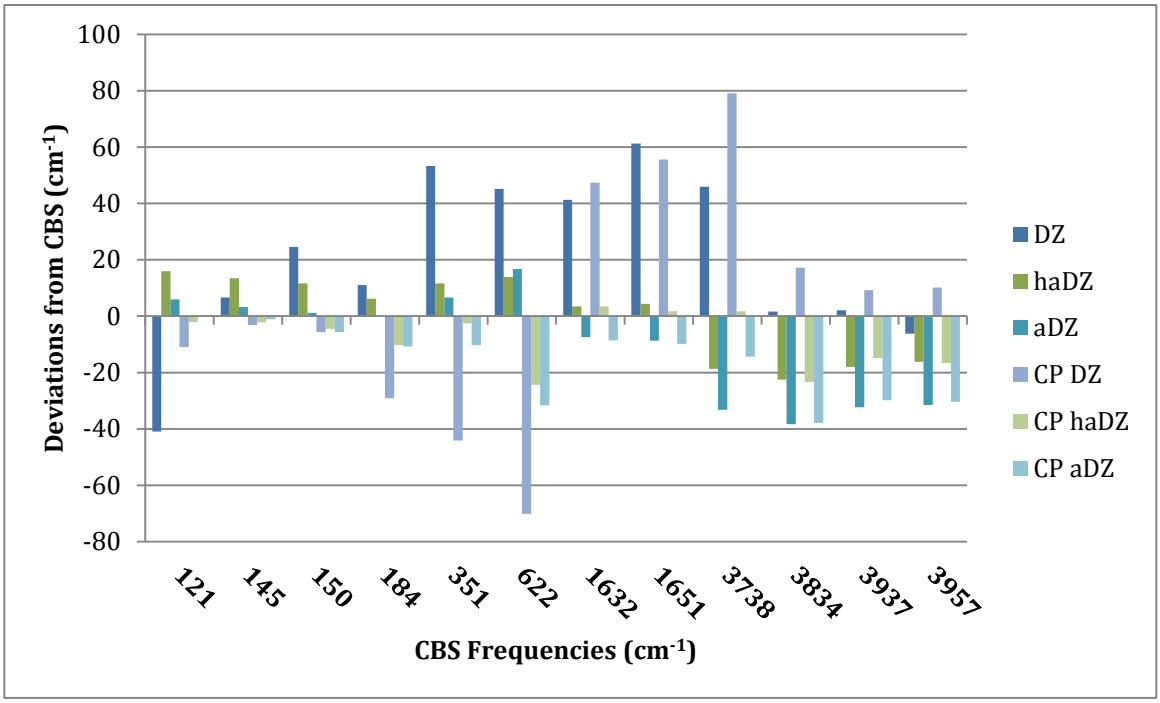

Fig. 10 Deviations from the estimated CBS limit harmonic vibrational frequencies in DZ quality basis sets with and without CP-correction procedures in $(H_2O)_2$

The deviations from the CBS estimate in the TZ, haTZ, and aTZ basis sets with and without CP-correction procedures are shown in *Figure 11*. The maximum absolute deviation from the CBS estimate in the non-CP corrected frequencies is 27 cm^{-1} and the maximum absolute deviation from the CBS estimate in the CP corrected frequencies is 48 cm^{-1} with the TZ basis set. With the haTZ basis set, the maximum absolute deviations shrink to 18 cm^{-1} in the non-CP corrected frequencies and 17 cm^{-1} in the CP-corrected frequencies. With the aTZ basis set, the maximum absolute deviations shrink to 22 cm^{-1} in the non-CP corrected frequencies and 20 cm^{-1} in the CP-corrected frequencies. When the size of the basis set is increased to the TZ quality level, CP-corrections appear beneficial when the haTZ and aTZ basis sets are used—and only decrease the deviation by 1-2 cm^{-1} over the non-CP corrected frequency deviations.

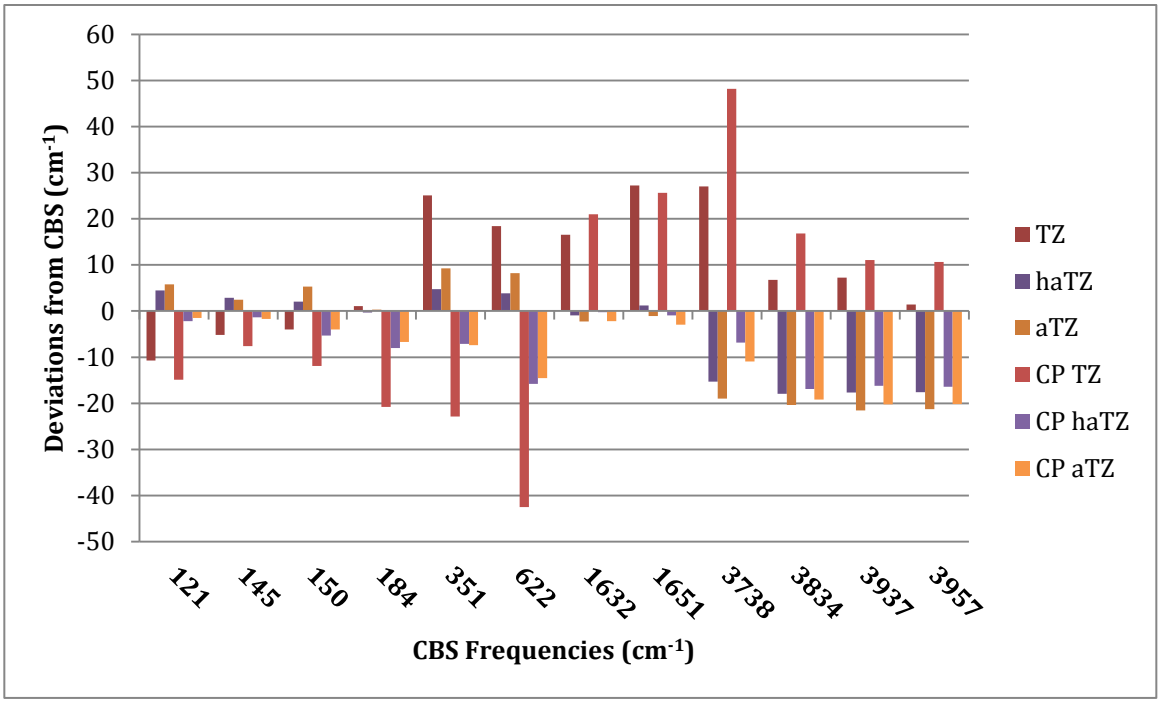

Fig. 11 Deviations from the estimated CBS limit harmonic vibrational frequencies in TZ quality basis sets with and without CP-correction procedures in $(H_2O)_2$

Figure 12 shows the deviations from the CBS estimate in the QZ, haQZ, and aQZ basis sets with and without CP-correction procedures. The maximum absolute deviation from the CBS estimate in the non-CP corrected frequencies is 18 cm^{-1} and the maximum absolute deviation from the CBS estimate in the CP corrected frequencies is 31 cm^{-1} with the QZ basis set. With the haQZ basis set, the maximum absolute deviations shrink to 3 cm^{-1} in the non-CP corrected frequencies and 7 cm^{-1} in the CP-corrected frequencies. With the aQZ basis set, the maximum absolute deviations shrink to 6 cm^{-1} in the non-CP corrected and the CP-corrected frequencies. These results indicate that in order to provide deviations from the CBS estimate of less than 10 cm^{-1}, a basis set of haQZ or larger needs to be utilized. These results also indicate that once a basis set of this size is used, the CP-corrections are not cost effective as the deviations in harmonic vibrational frequencies are already small.

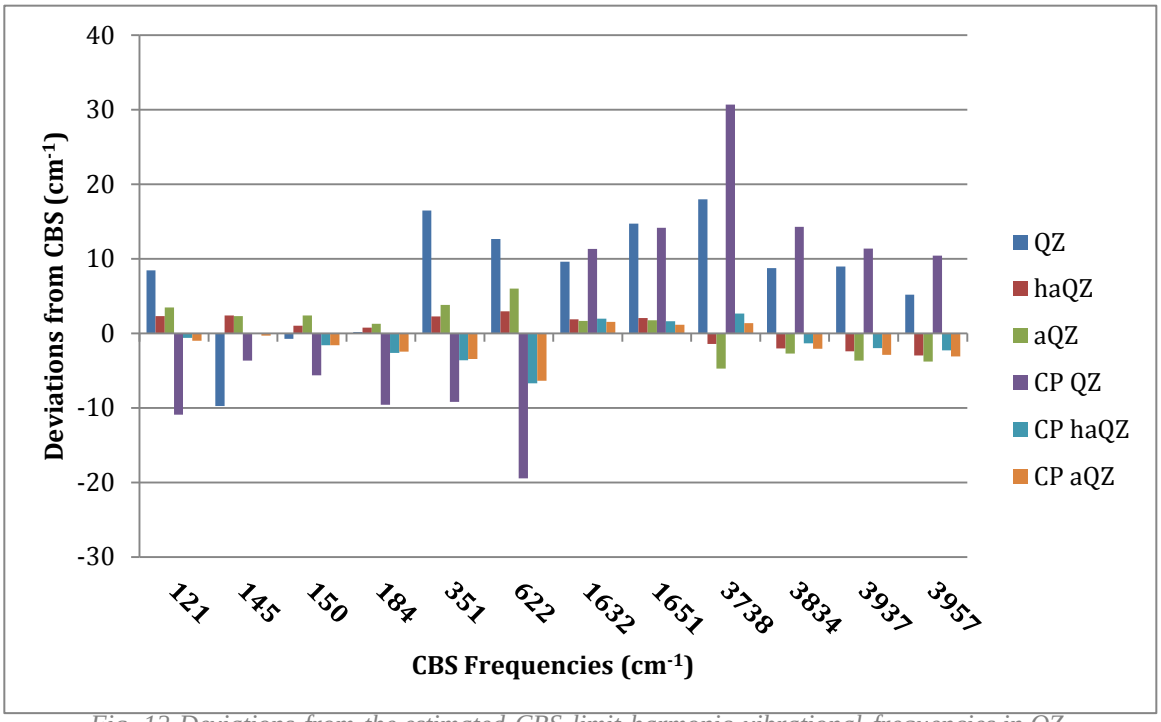

Fig. 12 Deviations from the estimated CBS limit harmonic vibrational frequencies in QZ quality basis sets with and without CP-correction procedures in $(H_2O)_2$

For both the $(HF)_2$ and $(H_2O)_2$ systems, a basis set of at least haQZ size is necessary in order to provide harmonic vibrational frequencies that deviate no more than 10 cm^{-1} from the estimated CBS limit. When a basis set of this size is used the CP-correction procedure is no longer cost effective.

4b. The Effects of Diffuse Functions

The results of the harmonic vibrational frequency computations in which specific diffuse functions were removed from the basis set suggest that some diffuse functions can be removed from basis sets without appreciably increasing the deviations from the estimated MP2 CBS limit. The haQZ and aQZ basis sets were used because—as stated above—a basis set of at least haQZ quality is capable of producing harmonic

vibrational frequencies that deviate no more than 10 cm^{-1} from the CBS estimate. From $(HF)_2$ in *Fig. 13*, the aQZ (f/g) denotes the harmonic vibrational frequencies acquired with the use of a custom basis set in which the set of *f* diffuse functions has been removed from the hydrogen atom and the set of *g* diffuse functions has been removed from the fluorine atom. The aQZ (df/fg) the use of a similar basis set in which the sets of *d* and *f* diffuse functions have been removed from the hydrogen atom and an *f* and *g* diffuse function have been removed from the fluorine atom. The haQZ (g) denotes the harmonic vibrational frequencies acquired with the use of a custom basis set in which a *g* diffuse function has been removed from the fluorine atom and so on with the (fg) denoting the removal of both an *f* and *g* diffuse function from the fluorine atom.

In *Fig. 14*, the aQZ (f/g), haQZ (g), and haQZ (fg) denote the same basis set modifications as they do in *Fig. 13*. Because the haQZ basis set frequencies are closer to the estimated CBS frequencies, all of the diffuse functions can be removed from the hydrogen atom in both dimer systems. In fact, doing so improves the harmonic vibrational frequencies with respect to the CBS limit. The results from both figures indicate that one set of *f* and *g* diffuse functions can be removed at from both the aQZ and haQZ basis set without causing a deviation from the CBS frequency values of more than ±10 cm^{-1}.

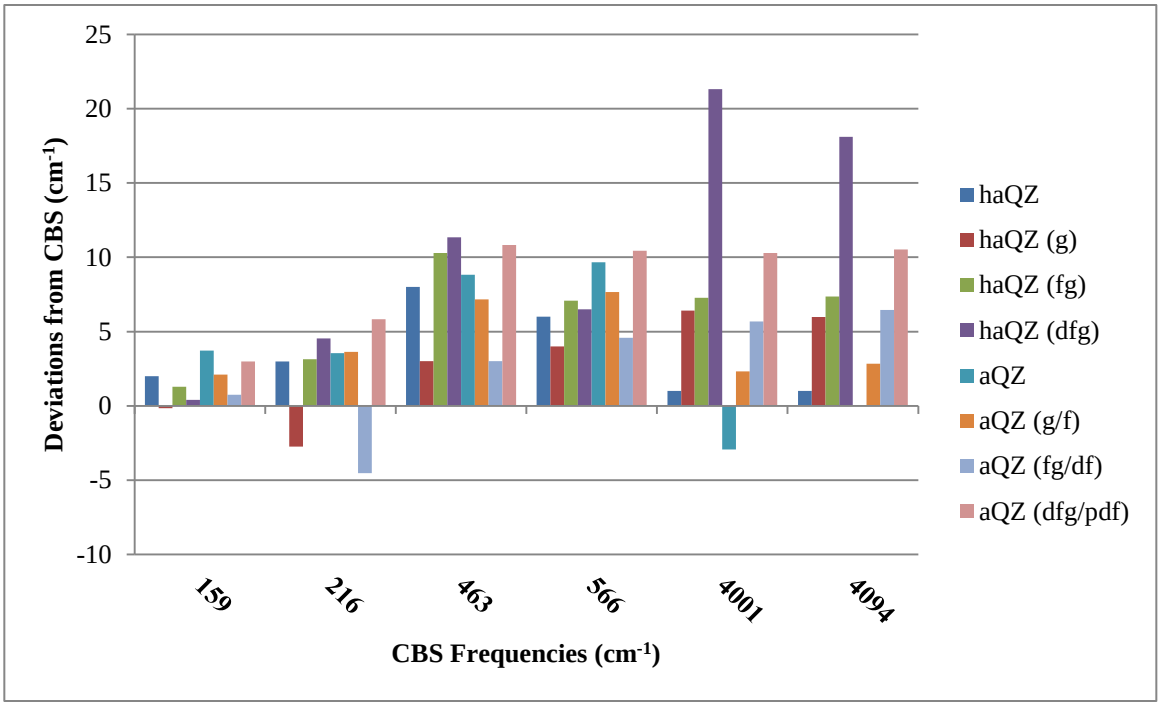

Fig. 13 The deviation from the CBS limit with respect to harmonic vibrational frequencies when diffuse functions are removed from the aQZ and haQZ basis set in $(HF)_2$

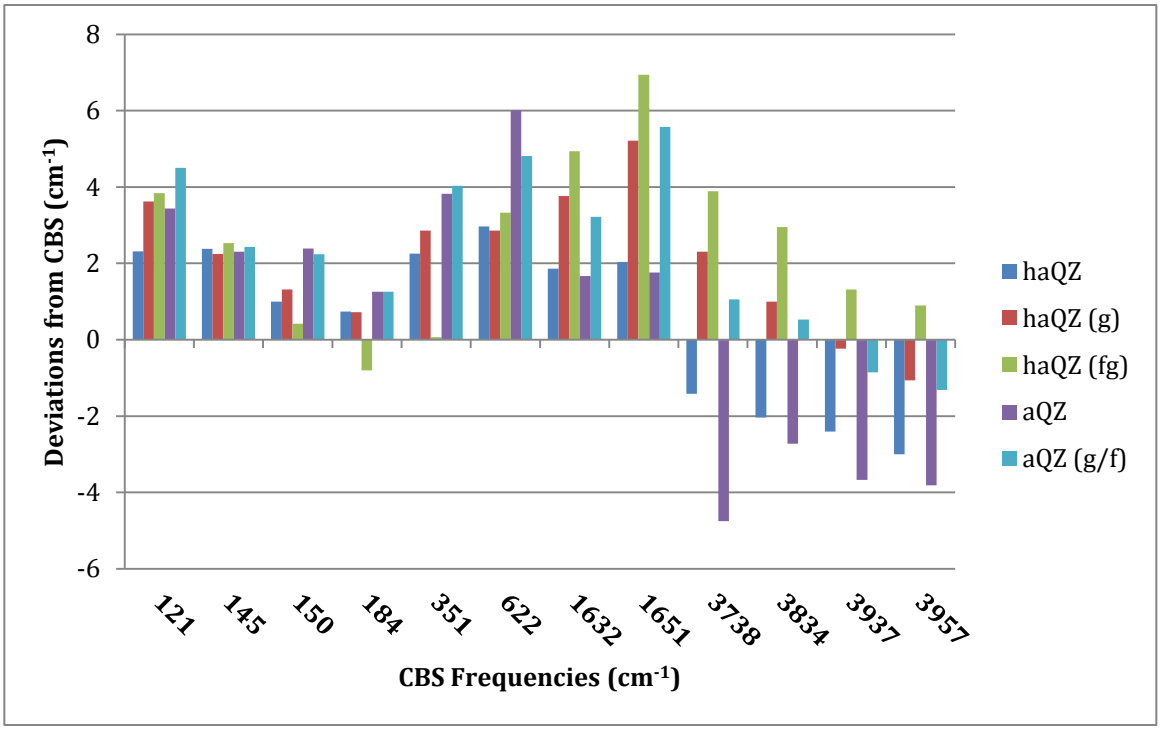

Fig. 14 The deviation from the CBS limit with respect to harmonic vibrational frequencies when diffuse functions are removed from the aQZ and haQZ basis set in $(H_2O)_2$

5. CONCLUSIONS

The results indicate that basis set selection has a significant effect on the harmonic vibrational frequencies of the $(HF)_2$ and $(H_2O)_2$ systems. For both systems, an MP2 CBS estimate was calculated using the average harmonic vibrational frequencies of the ha6Z, CP ha6Z, a6Z, and CP a6Z (when available). The computed harmonic vibrational frequencies from all other basis sets were compared to these CBS estimates. From examining the deviations from the estimated CBS frequencies, it is clear that in order to produce frequencies that do not deviate from the CBS limit by more than 10 cm^{-1}, an haQZ basis set or better must be utilized. Additionally, CP-corrected procedures are only somewhat effective when using smaller basis sets such as the DZ, TZ, haDZ, etc. CP-corrected procedures become ineffective when using an haQZ or bigger basis set. The additional time demands to finish computations that employ CP-correction procedures makes the use of CP-correction procedures a cost ineffective.

The harmonic vibrational frequencies acquired through the use of the custom aQZ and haQZ basis sets in which sets of *d*, *f*, and *g* diffuse functions have been removed from one or both of the atoms suggest that all of the diffuse functions can be removed from the hydrogen atom and that certain diffuse functions can be removed from the heavy atoms in both basis sets in order to save time without appreciably increasing the deviations from the

estimated MP2 CBS limit. Removing diffuse functions from the hydrogen atoms actually seems to decrease deviations from the CBS frequencies, and sets of diffuse *d* and *f* functions can be removed from the heavy atoms (oxygen and fluorine) without causing a frequency deviation of more than 10 cm^{-1} from the estimated MP2 CBS limit.

BIBLIOGRAPHY

1. Edward G. Hohenstein and C. David Sherrill. *WIREs Comput Mol Sci.*, 2(2):304–326, (2012). DOI: 10.1002/wcms.84
2. K. Rappé and E. R. Bernstein. *J. Phys. Chem. A.* 104(26):6117-6128, (2000). DOI: 10.1021/jp0008997
3. Kevin E. Riley, Michal Pitoňák, Petr Jurečka, and Pavel Hobza. *Chem. Rev.*, 110(9):5023-5063, (2010). DOI: 10.1021/cr1000173
4. Kevin E. Riley, James A. Platts, Jan Řezać, Pavel Hobza, and J. Grant Hill. *J. Phys. Chem. A.*, 116(16):4159−4169, (2012). DOI: 10.1021/jp211997b
5. J. A. Anderson and G. S. Tschumper. *J. Phys. Chem. A.*, 110(22):7268-7271 (2006). DOI: 10.1021/jp0613889
6. Sławomir J. Grabowski, W. Andrzej Sokalski, Edyta Dyguda, and Jerzy Leszczyński. *J. Phys. Chem. B.*, 110(13):6444-6446, (2006). DOI: 10.1021/jp0600817
7. Daniel Boese, Georg Jansen, Martin Torheyden, Sebastian Höfener and Wim Klopper. *Phys. Chem. Chem. Phys.*, 13(3):1230–1238, (2011). DOI: 10.1039/C0CP01493A
8. A. Daniel Boese. *J. Chem. Theory Comput.* 9(10):4403−4413, (2013). DOI: 10.1021/ct400558w
9. E. N. Baker and R. E. Hubbard. *Prog. Biophys. Mol. Biol.*, 44(2):97-179, (1984). DOI: 10.1016/0079-6107(84)90007-5
10. Bernd Kuhn, Peter Mohr, and Martin Stahl. *J. Med. Chem.*, 53(6):2601-2611, (2010). DOI: 10.1021/jm100087s
11. Young Choon Park and Jae Shin Lee. *Bull. Korean Chem. Soc.*, 28(3):386-390 (2007). Available online from: http://pdf.lookchem.com/pdf/32/4f21c91d-4426-444c-9190-897383dec7d1.pdf
12. Ahmed El Kerdawy, Christofer S. Tautermann, Timothy Clark, and Thomas Fox. *J. Chem. Inf. and Model.*, *53*(12):3262-3272, (2013). DOI: 10.1021/ci4006222
13. James U. Bowie., *Curr Opin Struct Biol*. 21(1): 42–49, (2011). DOI:10.1016/j.sbi.2010.10.003
14. Ian K. McDonald, Janet M. Thornton. *J. Mol. Biol.*, 238(5):777-793, (1994). DOI: 10.1006/jmbi.1994.1334
15. K.L. Copeland and G.S. Tschumper, *J. Chem. Theory Comput.*, 8(5):1646-1656 (2012). DOI: 10.1021/ct300132e
16. J. A. Anderson, B. W. Hopkins, J. L. Chapman, and G. S. Tschumper. *J. Mol. Struct. (THEOCHEM)*, 771:65-71 (2006). DOI:10.1016/j.theochem.2006.03.042
17. William D. Arnold and Eric Oldfield. *J. Am. Chem. Soc.*, 2000, 122 (51):12835–12841. DOI: 10.1021/ja0025705
18. Martin W. Feyereisen, David Feller, and David A. Dixon. *J. Phys. Chem.*, 100(8), 2993–2997, (1996). DOI: 10.1021/jp952860l
19. Howard Perlman. The USGS water science school. (2014). Available on: http://water.usgs.gov/edu/waterproperties.html
20. Xin Xu and William A. Goddard, III. *J. Phys. Chem. A, Vol.* 108(12):2305-2313, (2004) DOI: 10.1021/jp035869t
21. Teemu Salmi, Vesa Hänninen, Anna L. Garden, Henrik G. Kjaergaard, Jonathan Tennyson, and Lauri Halonen. *J. Phys. Chem. A.* 112(28):6305–6312, (2008). DOI: 10.1021/jp800754y
22. Joseph R. Lane. *J. Chem. Theory Comput.* 9(1):316−323, (2013). DOI: 10.1021/ct300832f
23. K. Kim and K. D. Jordan. *J. Phys. Chem.*, 98(40):10089–10094, (1994). DOI: 10.1021/j100091a024
24. Alex Shank, Yimin Wang, Alexey Kaledin, Bastiaan J. Braams, and Joel M. Bowman. *J. Chem. Phys*. 130(14):144314, (2009). DOI: 10.1063/1.3112403
25. Gregory S. Tschumper, Matthew L. Leininger, Brian C. Hoffman, Edward F. Valeev, Henry F. Schaefer et al. *J. Chem. Phys.*, 116(2):690, (2002). DOI: 10.1063/1.1408302
26. J.C. Howard and G.S. Tschumper, *WIREs Comput. Mol. Sci.*, published online, (2013). DOI: 10.1002/wcms.1168
27. D.M. Bates, J.R. Smith, T. Janowski and G.S. Tschumper, *J. Chem. Phys.*, 135:044123 (2011). DOI: 10.1063/1.3609922
28. Thomas Engel and Phillip Reid. *QuantumChemistry and Spectroscopy*. (3) 2012.
29. Martin Head-Gordon, John A. Pople, and Michael J. Frisch. *Chem. Phys. Lett.*, 153(6):503- 506, (1988). DOI: 10.1016/0009-2614(88):85250-3
30. Asger Halkier, Henrik Koch, Poul Jørgensen, Ove Christiansen, Ida M. Beck Nielsen, and Trygve Helgaker. *Theory. Chem. Acc.* 97(1-4), 150-157, (1997). Available online from: http://iqc.udg.es/articles/bsse/1997-tca-helgaker.pdf
31. S. F. Boys and F. Bernardi, Mol. Phys. 19(4):553 (1970). DOI: 10.1080/00268977000101561
32. Michael J. Frisch, Martin Head-Gordon and John A. Pople. Chem. Phys. Lett.,166(3):275-280, (1990). DOI: 10.1016/0009-2614(90)80029-D
33. Pavel Hobza, Ota Bludský and Sándor Suhai *Phys. Chem. Chem. Phys.*, 1:3073-3078 (1999). DOI: 10.1039/A902109D
34. Béla Paizs and Sándor Suhai. *J. Comp. Chem.*, 19(6):575-684. (1998). DOI: 10.1002/(SICI)1096-987X(19980430)19:6<575::AID-JCC1>3.0.CO;2-O
35. Frans B. van Duijneveldt, Jeanne G. C. M. van Duijneveldt-van de Rijdt, and Joop H. van Lenthe. *Chem. Rev.*, 94(7):1873-1885, (1994). DOI: 10.1021/cr00031a007
36. Lori A. Burns, Michael S. Marshall, and C. David Sherrill. *J. Chem. Theory Comput.*

10(1):49-57, (2014). DOI: 10.1021/ct400149j
37. Ł. M. Mentel and E. J. Baerends. *J. Chem. Theory Comput.* 10(1):252−267 (2014). DOI: 10.1021/ct400990u
38. C. David Sherill. (2010). Available on: http://vergil.chemistry.gatech.edu/notes/cp.pdf
39. Matthew Goldey, Anthony Dutoi and Martin Head-Gordon. *Phys. Chem. Chem. Phys.*, 15(38):15869-15875, (2013). DOI: 10.1039/c3cp51826d
40. Joachim Friedrich. *J. Chem. Theory Comput.* 8(5):1597−1607, (2012). DOI: 10.1021/ct200686h
41. David Feller. *J. Chem. Phys.* 96(8):6104-6115, (1992). DOI: 10.1063/1.462652
42. Narbe Mardirossian and Martin Head-Gordon. *J. Chem. Theory Comput.*, 9(10):4453−4461, (2013). DOI: 10.1021/ct400660j
43. Ernest R. Davidson and David Feller. *Chem. Rev.*, 86(4):681-696, (1986). DOI: 10.1021/cr00074a002
44. Henk Eshuisa and Filipp Furcheb *J. Chem. Phys.*, 136(8):084105, (2012). DOI: 10.1063/1.3687005
45. D.M. Bates, J.R. Smith and G.S. Tschumper, *J. Chem. Theory Comput.*, 7(9):2753-2760 (2011). DOI: 10.1021/ct200176t
46. László Turi and J. J. Dannenberg. *J. Phys. Chem.* 97(11):2488-2490, (1993). DOI: 10.1021/j100113a002
47. J. G. C. M. van Duijneveldtvan de Rijdt and F. B. van Duijneveldt. *J. Chem. Phys.* 97(7):5019, (1992). DOI: 10.1063/1.463856
48. Trygve Helgaker, Wim Klopper, Henrik Koch, and Jozef Noga. *J. Chem. Phys.* 106(23), 9639, (1997). DOI: 10.1063/1.473863
49. Andrew G. Taube and Rodney J. Bartlett. *Collect. Czech. Chem. Commun.*, 70(6):837- 850, (2005). DOI: 10.1135/cccc20050837
50. Joachim Friedrich, Eva Perlt, Martin Roatsch, Christian Spickermann, and Barbara Kirchner. *J. Chem. Theory Comput.* 7(4):843–851, (2011). DOI: 10.1021/ct100131

APPENDIX

APPENDIX 1A

Procedure for research using water as the example.

1. Create an opt.com file to run in Gaussian09
    a. Example input

```
%mem=100mb
%nprocs=2

#MP2/GEN 5d, 7f opt (tight) scf(conver-10)

Water dimer

0,1
O
H 1 R1
H 1 R2 2 A1
O 2 R3 1 A2 3 D1
H 4 R4 2 A3 3 D2
H 4 R4 2 A3 3 D3

R1=0.9706
R2=0.9641
R3=1.9439
R4=0.967
A1=101.8279
A2=175.0
A3=102.108
D1=180.0
D2=127.5678
D3=-127.5678

H
cc-PVDZ
***
O
cc-PVDZ
***
```

2. Type the g09 command to run the program
3. Using the optimized coordinates, create a frequency input file (freq.com)

a. Example input

```
%mem=3000mb
%nprocs=2

#MP2/GEN 5d, 7f  freq

Water dimer

0,1
O
H 1 R1
H 1 R2 2 A1
O 2 R3 1 A2 3 D1
H 4 R4 2 A3 3 D2
H 4 R4 2 A3 3 D3

R1=0.97057793
R2=0.96414257
R3=1.94396602
R4=0.9669895
A1=101.8282046
A2=172.66575051
A3=102.0969105
D1=180.0
D2=127.1813327
D3=-127.1813327

H
cc-PVDZ
***
O
cc-PVDZ
***
```

4. Run another optimization (opt.com) using cc-PVTZ in place of cc-PVDZ
5. Follow steps 2-3 with these coordinates to run a frequency computation
6. Run optimizations and frequencies in the same manner using aug-cc-PVXZ for both atoms
7. Run optimizations and frequencies in the same manner using cc-PVXZ for hydrogen (H) and aug-cc-PVXZ for oxygen (O)

    note that the X stands for D, T, Q, 5, and 6Z in that order

APPENDIX 1B

Actual Frequency results from research in table format

$(HF)_2$

Without CP-correction

	DZ	TZ	QZ	5Z	6Z	haDZ	haTZ	haQZ	ha5Z	ha6Z	aDZ	aTZ	aQZ	a5Z	a6Z
a'	131	160	164	162	160	160	158	160	160	159	158	159	162	162	160
a'	232	217	218	219	217	221	220	219	217	216	214	219	220	220	218
a”	513	491	476	470	466	474	479	470	465	463	471	474	471	472	468
a'	644	603	585	573	569	580	577	572	568	567	576	578	575	572	569
a'	4058	4053	4027	4011	4004	3955	3993	4003	4002	4001	3940	3989	3998	4001	4000
a'	4116	4129	4112	4102	4097	4054	4082	4095	4095	4094	4039	4081	4094	4094	4094

With CP-correction

a'	153	148	156	157	146	151	157	157	158	147	151	157	158	158
a'	205	203	211	214	206	210	214	215	215	205	209	214	215	215
a”	437	450	459	461	440	451	458	459	460	433	450	458	460	460
a'	517	532	553	560	541	552	560	562	563	537	550	560	563	563
a'	4096	4075	4039	4017	3985	4001	4008	4005	4002	3969	3999	4007	4002	4002
a'	4140	4143	4119	4105	4056	4084	4097	4095	4094	4043	4083	4096	4094	4094

$(H_2O)_2$

Without CP-correction

	DZ	TZ	QZ	5Z	6Z	haDZ	haTZ	haQZ	ha5Z	ha6Z	aDZ	aTZ	aQZ	a5Z	a6Z
a”	80	110	130	126	122	137	126	124	123	122	127	127	125	123	121
a'	151	139	135	145	146	158	148	147	146	146	148	147	147	146	143
a”	174	146	149	152	150	161	152	151	150	150	151	155	152	151	150
a'	195	185	184	185	184	190	184	185	184	184	184	184	185	185	185
a'	404	376	368	358	353	363	356	353	352	352	358	360	355	354	352
a”	667	640	635	628	624	636	626	625	624	623	639	630	628	626	624
a'	1673	1648	1641	1636	1634	1635	1631	1633	1633	1632	1624	1629	1633	1633	1632
a'	1713	1679	1666	1658	1655	1656	1653	1653	1652	1651	1643	1650	1653	1653	1651
a'	3784	3765	3756	3745	3740	3719	3722	3736	3738	3738	3704	3719	3733	3737	3737
a'	3836	3841	3843	3839	3836	3812	3816	3832	3835	3834	3796	3814	3832	3834	3834
a'	3939	3944	3946	3942	3939	3919	3919	3934	3937	3937	3904	3915	3933	3937	3937
a”	3950	3958	3962	3960	3958	3940	3939	3954	3957	3957	3925	3935	3953	3956	3956

With CP-correction

a”	110	106	110	116	119	119	119	121	121	121	121	120	120	120
a'	141	137	141	145	145	142	143	145	145	145	144	143	144	145
a”	144	138	144	146	148	145	144	148	149	149	144	146	148	149
a'	155	163	174	180	182	174	176	181	183	183	173	177	181	183
a'	307	328	342	348	350	349	344	348	349	350	341	344	348	349
a”	552	580	603	615	619	598	606	615	619	620	590	608	616	618
a'	1679	1653	1643	1637	1634	1635	1631	1634	1633	1632	1623	1629	1633	1633
a'	1707	1677	1665	1657	1654	1653	1650	1653	1652	1651	1641	1648	1652	1652
a'	3817	3786	3768	3751	3743	3739	3731	3740	3740	3739	3723	3727	3739	3740
a'	3851	3851	3849	3841	3837	3811	3817	3833	3835	3835	3796	3815	3832	3835
a'	3946	3948	3948	3943	3939	3922	3920	3935	3938	3937	3907	3917	3934	3937
a”	3967	3967	3967	3962	3959	3940	3940	3954	3957	3957	3926	3936	3953	3957